

Learn How to Perform a t-Test for Regression Slope in R

Authored by
Mohammed looti

November 16, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Learn How to Perform a t-Test for Regression Slope in R*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=2754>

In the foundational discipline of statistics, [linear regression](#) serves as an indispensable analytical technique. It is primarily utilized to establish and quantify the linear relationship between a [response variable](#) (dependent variable) and one or more [predictor variables](#) (independent variables). When conducting a simple linear regression, the main objective is twofold: to accurately predict an outcome based on changes in the predictor, and to precisely characterize the nature and strength of this relationship. This modeling process ultimately yields an estimated regression equation, which mathematically encapsulates the trend observed within the collected data.

The standard representation of this estimated linear regression equation is given by:

$$\hat{y} = b_0 + b_1x$$

In this equation, $\hat{y}$ denotes the predicted value of the [response variable](#), while x represents the value of the [predictor variable](#). The term b_0 is the [y-intercept](#), representing the predicted outcome when the predictor variable x is zero. Most critically, b_1 is the [slope coefficient](#). This coefficient is paramount in regression analysis, as it rigorously quantifies the average magnitude of change expected in the response variable for every single unit increase in the predictor variable.

Assessing the Significance of the Regression Slope

Once a [linear regression](#) model has been successfully fitted to the data, a critical subsequent step is to determine the authenticity of the observed relationship. We must ascertain whether the apparent link between the variables is a genuine underlying phenomenon or merely a byproduct of random sampling variability. This is achieved by formally assessing if the [slope coefficient](#), b_1 , is [statistically significant](#).

A finding of [statistical significance](#) for the slope implies that there is compelling evidence for a reliable, non-zero linear association between the predictor and response variables within the larger population from which the data was drawn. Conversely, if the slope is found to be non-significant, it suggests that the predictor variable does not contribute a meaningful linear impact on the response variable, potentially rendering the predictor useless for forecasting outcomes within this linear framework.

Therefore, executing a formal test for the significance of the slope is not optional; it is an essential component of correctly interpreting any [linear regression](#) model. To formally evaluate the significance of b_1 , we rely on [hypothesis testing](#), specifically employing a [t-test](#). This procedure enables us to draw robust [statistical inferences](#) about the population slope (denoted as β_1) based entirely on our sample data. The outcome of the t-test directly informs our decision regarding the inclusion or exclusion of the [predictor variable](#) in the finalized model.

Formulating Hypotheses and Calculating the t-Statistic

The process of performing the [t-test](#) for the [slope coefficient](#) requires the establishment of two mutually exclusive hypotheses concerning the true population slope (β_1):

Null Hypothesis (H0): $\beta_1 = 0$. This hypothesis posits that the true population slope is zero, meaning no linear relationship exists between the variables.

Alternative Hypothesis (HA): $\beta_1 \neq 0$. This hypothesis suggests that the true population slope is non-zero, thereby indicating the presence of a significant linear relationship.

To test these hypotheses, we must compute a [test statistic](#). This statistic quantifies the distance, measured in units of [standard errors](#), between our estimated sample slope (**b1**) and the hypothesized population slope (zero, under H0). The mathematical formula used for calculating the t-test statistic for the slope is straightforward:

$$t = b1 / se(b1)$$

In this formula:

b1 is the calculated [slope coefficient](#) derived directly from our sample data.

se(b1) represents the [standard error](#) of **b1**. The standard error is crucial as it reflects the precision of the slope estimate; a lower value indicates higher confidence in the estimated slope's reliability.

Upon calculating the [test statistic](#), we reference the [t-distribution](#). We typically use the corresponding [p-value](#) to facilitate the decision-making process regarding the [Null Hypothesis](#). The associated [degrees of freedom](#) for this specific t-test are calculated as **n-2**, where 'n' is the total number of observations. The subtraction of two accounts for the two parameters (the intercept and the slope) that are estimated in a simple linear regression model.

Interpreting the p-value and Making a Decision

The [p-value](#) is arguably the most vital output of [hypothesis testing](#). It defines the probability of observing a [test statistic](#) as extreme as, or even more extreme than, the one computed from our sample data, under the strict assumption that the [null hypothesis](#) (no relationship) is true. Simply put, it measures the likelihood of obtaining our results if no true linear relationship exists in the population.

To reach a conclusion, the calculated [p-value](#) is compared against a pre-established [significance level](#), conventionally denoted as α (alpha). Standard choices for α are usually 0.05 or 0.01. The decision rule is clearly defined:

If the [p-value](#) is **less than** α (e.g., $p < 0.05$), we must **reject the null hypothesis**. This outcome

provides sufficient evidence to conclude that the [slope coefficient](#) is truly [statistically significant](#) and distinct from zero. Consequently, we infer that a meaningful linear relationship exists between the predictor and response variables.

If the [p-value](#) is **greater than or equal to** α (e.g., $p \geq 0.05$), we **fail to reject the null hypothesis**. This implies that the data does not offer enough compelling evidence to establish that the [slope coefficient](#) is [statistically significant](#), suggesting the observed relationship could plausibly be attributed to random sampling variation.

The rejection of the [null hypothesis](#) confirms that the [predictor variable](#) genuinely contributes explanatory power to the variation observed in the [response variable](#) within the context of the fitted model.

Practical Application: Executing the t-Test in R

To reinforce these theoretical concepts, we will now implement a concrete example using the [R programming language](#). R is universally recognized as a robust, flexible, and essential environment for statistical computing and professional data visualization.

Imagine we are investigating 12 students to see if the hours they spent studying correlate linearly with their resulting final exam scores. Our primary question is whether the number of hours studied can significantly predict the final score. We begin by structuring this information into a [data frame](#) in R:

#create data frame

```
df <- data.frame(hours=c(1, 1, 2, 2, 3, 4, 5, 5, 5, 6, 6, 8),  
score=c(65, 67, 78, 75, 73, 84, 80, 76, 89, 91, 83, 82))
```

#view data frame

```
df
```

```
hours score
```

```
1 1 65
```

```
2 1 67
```

```
3 2 78
```

```
4 2 75
```

```
5 3 73
```

```
6 4 84
```

```
7 5 80
```

```
8 5 76
```

```
9 5 89
```

```
10 6 91
```

11 6 83

12 8 82

The resulting [data frame](#), named 'df', contains 'hours' (our [predictor variable](#)) and 'score' (our [response variable](#)). The next step is to fit a [simple linear regression](#) model using this data to test for the existence of a [statistically significant](#) linear relationship.

Fitting the Model and Analyzing the R Output

To fit the model in the [R programming language](#), we employ the powerful [lm\(\) function](#), short for "linear model." The model syntax is specified as response ~ predictor, pointing it to our 'df' [data frame](#). Immediately after fitting the model, using the `summary()` function provides a comprehensive [model summary](#), furnishing all the essential statistics required to conduct our formal [t-test](#) for the slope.

```
#fit simple linear regression model
```

```
fit <- lm(score ~ hours, data=df)
```

```
#view model summary
```

```
summary(fit)
```

Call:

```
lm(formula = score ~ hours, data = df)
```

Residuals:

Min 1Q Median 3Q Max

-7.398 -3.926 -1.139 4.972 7.713

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 67.7685 3.3757 20.075 2.07e-09 ***

hours 2.7037 0.7456 3.626 0.00464 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.479 on 10 degrees of freedom

Multiple R-squared: 0.568, Adjusted R-squared: 0.5248

F-statistic: 13.15 on 1 and 10 DF, p-value: 0.004641

Based on the output provided by `summary(fit)`, we can construct the specific estimated regression equation for predicting the student scores:

Exam score = 67.7685 + 2.7037(hours)

Focusing on the "Coefficients" table for the 'hours' [predictor variable](#), we extract the key metrics necessary for the t-test:

Estimate (b1): 2.7037. This is the estimated [slope coefficient](#). Interpretation: For every additional hour studied, the predicted exam score increases by approximately 2.70 points.

Std. Error (se(b1)): 0.7456. This value represents the [standard error](#) of the slope, indicating the expected variability of this estimate across different samples.

t value: 3.626. This is the calculated [t-test statistic](#), computed as $2.7037 / 0.7456$.

Pr(>|t|): 0.00464. This is the associated two-sided [p-value](#) for the test statistic.

While the primary focus is the t-test, the [model summary](#) also provides other valuable statistics, such as the [Residual standard error](#) (a measure of model fit), the [Multiple R-squared](#) (proportion of variance explained), and the overall model [F-statistic](#).

Final Conclusion and Statistical Inference

To finalize our hypothesis test, we focus on the calculated results for the 'hours' variable:

t-statistic: **3.626**

p-value: **0.00464**

By employing a customary [alpha level](#) (α) of 0.05, we compare our obtained [p-value](#) (0.00464) to this threshold. Since 0.00464 is definitively less than 0.05, we possess strong statistical justification to **reject the [null hypothesis](#)** ($H_0: \beta_1 = 0$).

Our resulting inference is clear: the [slope coefficient](#) quantifying the relationship between hours studied and exam score is [statistically significant](#). This evidence strongly suggests that the linear relationship observed in our sample is genuine and highly unlikely to be purely random. Therefore, we can confidently assert that, according to this analysis, dedicating more hours to studying is significantly associated with achieving higher final exam scores.

Beyond the t-Test: Expanding Analysis in R

Mastering the execution and interpretation of the [t-test](#) for the [slope coefficient](#) is foundational to performing competent [linear regression](#) analysis. The [R programming language](#) not only facilitates this core analysis efficiently but also offers an expansive ecosystem for sophisticated statistical diagnostics.

Beyond simply testing significance, analysts routinely use R to perform crucial follow-up tasks, such as validating core model assumptions, generating visual plots of residuals to check for

linearity and homogeneity of variance, and calculating [confidence intervals](#) for the slope estimate, which provide a range of plausible values for the true population slope.

For those aspiring to advanced analytical proficiency, exploring R's extensive capabilities for advanced statistical modeling and diagnostics is highly recommended. Utilizing the numerous high-quality online tutorials and official documentation available will significantly enhance your ability to handle more complex statistical challenges and interpret multivariate models with precision.